RESEARCH ARTICLE

Automated detection of immune effector cell-associated neurotoxicity syndrome via quantitative EEG

Christine A. Eckhardt[1,2,3,4] [iD], Haoqi Sun[1,2,3], Preeti Malik[1,2,3], Syed Quadri[1,2,3],
Marcos Santana Firme[1,2,3], Daniel K. Jones[1,2,3,5], Meike van Sleuwen[1,2,3], Aayushee Jain[1,3],
Ziwei Fan[1,2,3], Jin Jing[1,2,3], Wendong Ge[1,2,3], Husain H. Danish[1,2,4], Caron A. Jacobson[6],
Daniel B. Rubin[1,2], Eyal Y. Kimchi[1,2], Sydney S. Cash[1,2], Matthew J. Frigault[2,7,8], Jong Woo Lee[2,4] [iD],
Jorg Dietrich[1,2,9] & M. Brandon Westover[1,2,3,9]

[1]Department of Neurology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA
[2]Harvard Medical School, Boston, Massachusetts 02115, USA
[3]Clinical Data Animation Center, Massachusetts General Hospital, Boston, Massachusetts 02114, USA
[4]Department of Neurology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA
[5]Brigham Young University, Provo, Utah 84602, USA
[6]Dana Farber Cancer Institute, Boston, Massachusetts 02115, USA
[7]Cellular Immunotherapy Program, Massachusetts General Hospital Cancer Center, Boston, Massachusetts 02114, USA
[8]Blood and Marrow Transplant Program, Massachusetts General Hospital, Boston, Massachusetts 02114, USA
[9]Massachusetts General Hospital Cancer Center for Brain Health, Boston, Massachusetts 02114, USA

Correspondence
Christine Eckhardt, Department of Neurology, Massachusetts General Hospital (MGH), 15 Parkman St., Suite 835, Boston, MA, USA. Tel: 908-403-4541; Fax: 580-297-9169; E-mail: ceckhardt@partners.org

Abstract

Objective: To develop an automated, physiologic metric of immune effector cell-associated neurotoxicity syndrome among patients undergoing chimeric antigen receptor-T cell therapy. **Methods**: We conducted a retrospective observational cohort study from 2016 to 2020 at two tertiary care centers among patients receiving chimeric antigen receptor-T cell therapy with a CD19 or B-cell maturation antigen ligand. We determined the daily neurotoxicity grade for each patient during EEG monitoring via chart review and extracted clinical variables and outcomes from the electronic health records. Using quantitative EEG features, we developed a machine learning model to detect the presence and severity of neurotoxicity, known as the EEG immune effector cell-associated neurotoxicity syndrome score. **Results**: The EEG immune effector cell-associated neurotoxicity syndrome score significantly correlated with the grade of neurotoxicity with a median Spearman's R^2 of 0.69 (95% CI of 0.59–0.77). The mean area under receiving operator curve was greater than 0.85 for each binary discrimination level. The score also showed significant correlations with maximum ferritin (R^2 0.24, $p = 0.008$), minimum platelets (R^2 –0.29, $p = 0.001$), and dexamethasone usage (R^2 0.42, $p < 0.0001$). The score significantly correlated with duration of neurotoxicity (R^2 0.31, $p < 0.0001$). **Interpretation**: The EEG immune effector cell-associated neurotoxicity syndrome score possesses high criterion, construct, and predictive validity, which substantiates its use as a physiologic method to detect the presence and severity of neurotoxicity among patients undergoing chimeric antigen receptor T-cell therapy.

Introduction

Chimeric antigen receptor-T (CAR-T) cell therapy is a powerful treatment for refractory and relapsed hematologic malignancies. However, a substantial proportion of patients experience toxicity during treatment, in the forms of cytokine release syndrome (CRS) and/or immune effector cell-associated neurotoxicity syndrome (ICANS).[1] Occurring in 20–70% of patients,[2–5] ICANS is a clinical syndrome characterized by decreased arousal and focal neurologic findings that can cause cerebral edema and death.[6–9]

Early and accurate diagnosis of ICANS is paramount to guide management of both CRS and ICANS. CRS often precedes and persists concurrently with ICANS, but tocilizumab, a common treatment for CRS, may worsen ICANS.[10–12] Dexamethasone is the mainstay of treatment for ICANS but may diminish the efficacy of CAR-T against the underlying malignancy with long-term use.[13] As a result, there is no consensus on the optimal daily dosage of dexamethasone for ICANS. These considerations highlight the need for accurate diagnostics to inform management decisions.

Despite the high prevalence of ICANS, an objective and easily repeatable biomarker of ICANS has not been established. Serum inflammatory markers, such as CRP, LDH, and ferritin, have been associated with ICANS, but lack specificity given their elevation in CRS.[8,10] CSF profiles may reveal high levels of IL-6, IL-8, MCP1, and IP10 as well as a non-specific pleocytosis,[6] but invasive testing limits their use. Neuroimaging, such as CT and MRI, is typically normal except in severe ICANS leading to cerebral edema.[1,14,15] Multivariable models have shown promise in assessing the overall risk of ICANS, but do not respond to dynamic changes in the grade of ICANS on the timeframe of hours.[16,17] As a result, diagnosis of ICANS relies upon a standardized bedside neurological exam[7] that remains subjective and resource intensive.

EEG has the potential to serve as a physiologic biomarker of ICANS. Patients experiencing ICANS demonstrate profound qualitative changes in EEG, including increased delta and theta activity and generalized periodic discharges (GPDs), with the severity of certain features correlating with worsening ICANS.[18,19] However, small sample size and qualitative methods have limited the application of these findings to clinical practice. Nevertheless, prior studies have substantiated the use of EEG as a biomarker in disease states similar to ICANS, such as toxic-metabolic encephalopathy.[20–22]

Here, we present the EEG Immune effector cell associated neurotoxicity syndrome (EICANS) score, a quantitative method to determine the presence and grade of ICANS from EEG data through machine learning techniques. We demonstrate a strong correlation between ICANS grade and EICANS, demonstrating construct validity, and determine which EEG features are associated with worsening ICANS. We also show prominent associations between key clinical variables and outcomes and EICANS, which corroborates the construct and predictive validity of our model.

Methods

Study setting and participants

We conducted a multi-center, retrospective observational cohort study of adult inpatients who underwent long-term EEG monitoring during hospitalization for CD19 or B-cell maturation antigen (BCMA) targeted CAR-T cell therapy at Massachusetts General Hospital (MGH) and Brigham and Women's Hospital (BWH) from May 2016 to November 2020. The study was performed under a protocol approved by the Institutional Review Board using a waiver of written informed consent.

Clinical data

Generation of daily ICANS grade

Given changes in standardization of ICANS assessment from 2016 to 2019, for each day of hospitalization, the ICANS grade was determined through chart review by two graders according to ASTCT guidelines.[7]

Statistical analysis of patient characteristics

Quantitative data regarding patient characteristics were obtained via query of electronic health records. Data was compared between those with mild to moderate ICANS (≤2) and severe ICANS (>2), because these thresholds often have clinical implications for management decisions, such as ICU transfer.[1] Quantitative data are reported as medians (interquartile range) and compared using Mann–Whitney U tests. Categorical data are reported as n = counts (percent) and compared with chi-square tests. Significance level for these tests was defined as $p < 0.05$.

EEG dataset

Subjects were identified via chart review of patients who had a clinical EEG in the medical record and had undergone CAR-T cell therapy from 2016 to 2020. Samples were excluded from the dataset for the following reasons: (1) EEG prior to CAR-T cell infusion with none after CAR-T (2) EEG files could not be located within the server or contained excessive artifact (3) clinical exam could not be collected due to sedation requirements (4) EEG occurred greater than 30 days after infusion and neurological deficits were attributed to other causes by a neurologist. A maximum of five hospital days with EEG were utilized per patient to prevent overweighting of a test fold by a single patient. For those patients with greater than five EEGs, EEGs were selected semi-randomly to retain a similar distribution of ICANS grades to the original EEG files for that patient.

All EEGs were obtained due to clinical concern for ICANS; therefore, as a control dataset we used age and sex matched EEGs that were previously interpreted as normal by neurophysiologist physicians from BWH and

MGH from 2019 to 2021. For each EEG with a normal impression, the ICANS level was assigned to 0.

EEG recording, processing, and feature extraction

EEGs were collected with Ag/AgCL scalp electrodes with the standard international 10–20 system placed by EEG technicians. Tracings were re-sampled to 200 Hz, re-referenced to a bipolar montage, normalized to zero mean, notch filtered at 60 Hz, and bandpass filtered (0.5 Hz to 40 Hz).

Given the absence of a timestamp for each daily ICANS grade, an EEG block was selected to best avoid sleep (which could contain features considered normal during sleep but abnormal while awake, such as slowing). The mean alpha band power (8–12 Hz) across the four frontal leads was calculated for each 10-min EEG block every 5 min. To determine which 10-min block of EEG to utilize from 24 h of continuous EEG, the EEG block with the maximum alpha band power across the 24 h was selected, provided that this block contained 10% or fewer segments with artifact. The same algorithm was utilized to select a portion of each routine EEG with normal impressions.

After pre-processing, the EEG was divided into 10-sec segments with 8-sec of overlap (i.e., a segment every 2 sec, 10,800 segments) and any segments with artifact were flagged. EEG artifact were defined based on the presence of any of the following in any of the EEG leads: (1) maximum amplitude greater than 1000 μV (2) presence of NaN, representing when the EEG leads were disconnected. Further artifact detection methods were tested (standard deviation less than 0.2 μV and overly fast changes of greater than 900 uV in 0.1 s) in a subset of EEGs but did not yield appreciably different artifact detection rates, so were omitted to optimize for computational efficiency. All segments with artifact were removed, and EEG blocks with less than 10% artifact were truncated, such that all EEG blocks were 9 min in length.

Feature extraction & model training

For each 10-sec epoch, we calculated 94 total features, with 89 features in the time and frequency domains from the four frontal leads and 5 features derived from 16 bipolar leads with a convolutional neural network (CNN)-based model of interictal spectrum features (Table 2). Shoorangiz et al. (2020) provides an in-depth description of the majority of the time and frequency domain features.[23] With the exception of coherence features between the four frontal leads, each time and frequency-based feature was extracted for each frontal lead, then averaged across the frontal leads. For all time and frequency domain features, the mean

across all 10-sec epochs (265 total epochs) was taken to yield a single value for each feature per EEG file. For the CNN-derived features, the CNN model generated the probability of seizures, GPD, GRDA, LPD, LRDA, and Non-IIIC patterns for each epoch, summing to one for each epoch. The median across all epochs was taken to produce the final value for each CNN-derived feature. The seizure feature was ultimately excluded from the final model to reduce collinearity (given those six features sum to one), as this feature showed the lowest Spearman's correlation to ICANS.

We selected the top 20 features based on the training data for each fold using the F-statistic correlation quotient (FCQ) variant of the Maximum Relevance-Minimum Redundancy (MRMR) algorithm to reduce collinearity among the features, modified to use Spearman's instead of Pearson's correlation.[24] Hyperparameters were tuned through inner five-fold cross validation (within the training set), and features reduced through LASSO regularization.

Feature extraction and model training was performed using Python (v = 3.7.0). The periodic and aperiodic components of the physiological power spectrum were extracted utilizing the fitting oscillations and one over f (fooof) module (v = 1.0.0). A subset of univariate features ($n = 6$) were computed using the MNE-features module (v = 0.2). We performed model training with sci-kit learn modules (v 1.0.1). Scripts for the FCQ variant of MRMR were modified from the mrmr_selection module (v 0.2.2).

Learning-to-rank model training and evaluation

We utilized a machine learning model known as Learning-to-Rank (LTR) that transforms an ordinal classification problem into binary classification (logistic regression).[21,25] The LTR model predicts for each pair of patients A and B, whether patient A has more severe ICANS than B. Pairs with equivalent ICANS grades are excluded during training. The model accepts EEG features for each sample as input and outputs a score that correlates with the ICANS grade. For each EEG, the model generates a value for the decision function of the binary classifier (logistic regression), which we used to calculate the probability for each grade of ICANS (0–4). The final EICANS score for each file was the sum of these probabilities weighted by the ICANS grade (0–4) associated with each probability.

We trained the model with nested five-fold cross validation, which splits the data into training (80%) and test (20%) sets five times. The ICANS grades were stratified so that all test sets contained a similar distribution of ICANS grades, and data from the same patient were never split across training and test folds.

All reported performance metrics reflect the aggregated test sets of the five outer folds with strict separation of training and testing data. We utilized Spearman's R^2 and area under the receiver operator curve (AUC) to discriminate between binary severity levels of ICANS (i.e., ICANS ≤ 2 vs. >2). We generated confidence intervals (2.5% & 97.5% percentile as upper and lower bounds) and median coefficient values through bootstrapping 1000 times. The significance level for all tests was set at $p < 0.05$.

Association of ICANS with clinical variables and outcomes

To assess the clinical relevance of EICANS, we calculated the Spearman's correlation between EICANS and clinical variables, with *p*-values determined from a *t*-test with two degrees of freedom. Clinical variables, including medications, lab values, vital signs, and duration of hospitalization were obtained via query of electronic health records. Medications included antipsychotics (binary administration, IV, or PO formulation), dexamethasone (mg, IV, or PO), and tocilizumab dosage (# of doses) on the day of EICANS, as well as total dosage of antipsychotics (# of days), dexamethasone (total mg), and tocilizumab (total # of doses) over the course of hospitalization.

Results

Patient characteristics

136 patients who had undergone EEG monitoring and received CART-cell therapy were identified, of which 13 were excluded due to (1) EEG performed prior to, but not after CAR-T cell therapy ($n = 1$) (2) inability to locate raw EEG file or extract at least 9 min of continuous, artifact free EEG ($n = 9$) 4) inability to obtain exam consistently off of sedation ($n = 2$) and 4) encephalopathy attributed to cause other than ICANS in patient >30 days from infusion ($n = 1$). Of these 123 patients, 45 (36.6%) experienced mild to moderate ICANS (≤ 2) and 78 (63.4%) exhibited severe ICANS (>2) during hospitalization (Table 1). Three patients did not meet our criteria for ICANS on any day of hospitalization. Among patients who developed ICANS, the median time to onset was 6 days (IQR 4–8), similar to other studies.[1,8] Median duration of ICANS was higher among those with severe ICANS (13 days, IQR 9–19) as compared to mild to moderate ICANS (5, IQR 3–8, $p < 0.0001$), as was length of stay (median 24 vs. 20 days, $p = 0.0002$)). At 1-year post discharge, a higher but non-significant proportion of patients with severe as compared to mild to moderate ICANS were deceased (33 vs. 10, $p = 0.07$), with 23 patients with unknown status at 1 year. There was no significant difference in the distribution of sex, race, ethnicity, or malignancy between those with mild to moderate vs. severe ICANS. All patients were receiving an anti-seizure medication at the time of the EEG, with the majority (99.2%) on levetiracetam, one patient on lacosamide (3/5 EEG samples), and one patient on phenytoin (4/4 EEG samples). 3.25% (4/123) of patients received propofol on the day of EEG and 8.94% (11/123) of patients received a dose of lorazepam (10/123) or midazolam (1/123) on the day of EEG.

EEG dataset

From these 123 patients, we extracted one EEG sample per day with a maximum of 5 EEG samples per patient, totaling 286 EEG samples (median # samples per patient of 1 (IQR 1)). 184 (64.3%) EEG samples were collected during mild to moderate ICANS and 102 (35.7%) EEG samples during severe ICANS with a median ICANS of 2 (IQR 2) excluding control cases. All non-control EEGs were obtained due to concern for ICANS, indicating some degree of encephalopathy likely not captured by the day's ICANS even if graded as 0. The mean day of EEG sample relative to CAR-T cell infusion date was 11.05 ($\pm$ SD 6.25), and the mean day of EEG sample relative to first day of ICANS was 5.00 ($\pm$SD 5.28). As controls, the dataset was supplemented with 123 age and gender matched normal EEGs, resulting in 409 EEG samples utilized in the LTR model. EEGs from patients with differing grades of ICANS showed substantial qualitative differences, including within the same patient with ICANS 1 as compared to grade 2 (Fig. 1).

Criterion validity of EICANS, a physiological assessment of ICANS grade

For each EEG sample, we utilized the LTR model to produce a physiological assessment of ICANS grade: the Electroencephalographic ICANS (EICANS) score. EICANS scores possessed high criterion validity for ICANS, as shown by a significant correlation with the clinical ICANS grade from the same hospital day with a median Spearman's R^2 of 0.69 (95% CI of 0.59–0.77) on the aggregated test sets (Fig. 2A). In addition, for each binary discrimination level (ICANS $\leq x$ vs $> x$, for $x = 1$, 2, and 3), the mean area under the curve across all bootstraps was greater than 0.85 (Fig. 2B), although there was greater variability for discrimination at ICANS grade ≤ 3 vs. > 3 likely due to the low number of samples with ICANS >3 ($n = 7$). The receiver operating curve for discrimination of ICANS ≤ 2 vs. >2 for each bootstrap and the mean is shown in Fig. 2 (inset), with sensitivity of 85% and

	Total (n = 123)	ICANS ≤2 (n = 45)	ICANS >2 (n = 78)	p-value
Quantitative data[1,2]: unique subjects				
Age, years: Median (IQR)	64 (54–69)	62 (58–68)	65.5 (52–70)	
Length of stay (days): Median (IQR)	21 (17–30)	20 (16–27)	24 (18–32.5)	0.0002*
Duration ICANS >0: Median (IQR)	10 (5–17)	5 (3–8)	13 (9–19)	<0.0001*
Categorical data[1,3]: unique subjects				
Sex[4]				0.815
Male: n (%)	79 (64.2%)	30 (66.6%)	49 (62.8%)	
Female: n (%)	44 (35.7%)	15 (33.3%)	29 (37.2%)	
Race[4]				0.463
Asian: n (%)	4 (3.3%)	0 (0%)	4 (5.1%)	
Black: n (%)	2 (1.7%)	1 (2.3%)	1 (1.3%)	
White: n (%)	111 (90.2%)	42 (95.5%)	69 (88.5%)	
Other or Unknown: n (%)	6 (4.8%)	2 (4.4%)	4 (5.1%)	
Ethnicity[4]				0.418
Hispanic: n (%)	3 (2.4%)	2 (4.4%)	1 (1.3%)	
Non-Hispanic: n (%)	114 (92.7%)	40 (88.8%)	74 (94.9%)	
Unavailable: n (%)	6 (4.9%)	3 (6.6%)	3 (4.0%)	
Malignancy[4]				0.417
DLBCL: n (%)	106 (86.2%)	42 (95.5%)	64 (83.1%)	
PMBCL: n (%)	5 (4.1%)	0 (0%)	5 (6.5%)	
MCL: n (%)	2 (1.6%)	0 (0%)	3 (3.9%)	
FL: n (%)	4 (3.3%)	2 (4.5%)	2 (2.6%)	
MZL: n (%)	1 (0.8%)	0 (0%)	1 (1.3%)	
B-ALL: n (%)	2 (1.6%)	0 (0%)	2 (2.6%)	
Aggressive: n (%)	117 (95.1%)	43 (95.6%)	74 (94.9%)	
Indolent: n (%)	6 (4.9%)	2 (4.4%)	4 (5.1%)	
One-year post discharge[4]				0.070
Deceased: n (%)	43 (35.0%)	10 (22.2%)	33 (42.3%)	
Alive: n (%)	57 (46.3%)	35 (77.8%)	44 (56.4%)	
Unknown: n (%)	23 (18.7%)	9 (20.0%)	14 (17.9%)	
Anti-seizure medications (ASM)[4]				
On any ASM: n (%)	123 (100%)	45 (100%)	78 (100%)	
Keppra: n (%)	122 (99.2%)	44 (97.8%)	78 (100.0%)	0.780
Lacosamide: n (%)	3 (0.02%)	0 (0.0%)	3 (0.04%)	0.468
Phenytoin: n (%)	1 (0.01%)	1 (0.02%)	0 (0.0%)	0.780
Lorazepam: n (%)	10 (0.08%)	4 (0.09%)	6 (0.08%)	0.914
Midazolam: n (%)	1 (0.01%)	0 (0.0%)	1 (0.01%)	0.780
Anti-seizure sedation[4]				
Propofol: n (%)	4 (0.03%)	0 (0.0%)	4 (0.05%)	0.310
Continuous midazolam: n (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	

Table 1. Patient Characteristics according to ICANS grade ≤2 as compared to >2.

[1]Dataset includes 123 individual subjects who received CAR-T cell therapy.

[2]Quantitative data are shown as medians (IQR) or means (SD) and compared using Mann–Whitney U tests. The significance level for all tests was set at $p < 0.05$, indicated by *.

[3]Categorical data are shown as n = counts (percent) and compared using χ2 tests. The significance level for all tests was set at $p < 0.05$, indicated by *.

[4]Demographic data (age, sex, race, and ethnicity), subtype of malignancy, and death at 1-year post discharge are based on information extracted from the electronic health record.

specificity 82% at the optimal cutoff point. The mean absolute error was higher for ICANS >2 (0.86) than ICANS ≤2 (0.69), and there was a significant Spearman's correlation (0.26, $p < 0.0001$ between absolute error and maximum ICANS during hospitalization.

EEG features that predict ICANS

We examined the importance of individual quantitative EEG features in predicting ICANS. All features in the model with a median weight greater than or equal to 0.1

Table 2. Characteristics of extracted quantitative EEG features utilized in model.

Domain	Feature	Number	Remark
Time	Standard deviation, variance, mean absolute gradient, zero crossing rate, Hjorth mobility, Hjorth complexity, Hurst exponent (60 s & whole sample), skew, kurtosis, Higuchi fractal dimension, Katz fractal dimension	13	Mean across all 10-second epochs, generated from four frontal leads
Frequency	Mean power, spectral entropy, spectral edge frequencies: SEF95 and SEF5	4	Mean across all 10-second epochs, generated from four frontal leads
	Power Spectral Density (PSD) of different frequency bands and band-ratios, calculated for the PSD value in dB, with std. Calculated separately for different frequency bands and the band-ratios (e.g., PSD delta/ PSD alpha)	21 × 2 = 42	*Delta*: 0.5–2, 0.5–4, 2–4 Hz *Theta*: 4–6 Hz, 4–8, 6–8 Hz *Alpha*: 8–10, 8–12, 10–12 Hz *Beta*: 12–20 Hz All: 0.5–20 Hz Ratios: theta/delta, alpha/delta, alpha/theta, beta/ delta, beta/theta, beta/alpha Mean across all 10-second epochs, generated from four frontal leads
	Coherence for different frequency bands	24	Alpha (8–12 Hz), beta (12–20 Hz), theta (4–8 Hz), and delta (0.5–4 Hz). Mean across all 10-second epochs, generated from comparisons between each of four frontal leads
	FOOOF parameters: max amplitude, max frequency, max bandwidth, number of peaks, broadband offset, and exponent of aperiodic fit	6	FOOOF parameterizes neural power spectra Mean across all 10-second epochs, generated from four frontal leads
CNN derived	GPD, GRDA, LPD, LRDA, Non-IIC	5	Median across all 10-second epochs, generated from 18 leads
Total number of features per EEG		94	

($n = 12$) are visualized in Figure 3. Generalized periodic discharges (GPDs), generalized rhythmic delta activity (GRDA), and bandpower (4–6 Hz) were the strongest positive predictive features, while coherence of delta activity between leads Fp2-F8 and F7-F8 and the Hurst exponent were strong negative predictive features. The Non-IIC spectrum feature, which denotes EEG without evidence of potential epileptiform activity, was a negative predictor of EICANS.

Association of EICANS with clinical variables and outcome

We investigated the construct and predictive validity of EICANS scores through Spearman's correlations with clinical outcomes of interest (Fig. 4). The maximum EICANS per patient correlated with inflammatory markers previously associated with ICANS, including a significant correlation with the maximum serum ferritin (R^2 0.24, $p = 0.008$) and minimum serum platelets (R^2 − 0.29, $p = 0.001$). The maximum EICANS also significantly correlated with dexamethasone dosage on the day of EICANS (R^2 0.37, $p < 0.0001$) and total dexamethasone (R^2 0.42, $p < 0.0001$) and antipsychotics (R^2 0.19,

$p = 0.04$) during hospitalization. EICANS was associated with the duration of ICANS during hospitalization (R^2 0.31, $p = 0.0004$) but was not correlated with death by 1 year post discharge (R^2 0.01, $p = 0.84$). Maximum ICANS during EEG recording did not significantly correlate with death at 1 year post discharge (R^2 0.12, $p = 0.14$).

Discussion

In this retrospective study, we developed the EICANS score, a physiologic method to assess ICANS through EEG. Our demonstration of criterion, construct, and predictive validity establishes EICANS as a viable metric of ICANS that quantifies the presence and severity of ICANS among patients undergoing CAR-T cell therapy.

The correlation of EICANS and ICANS substantiates the criterion validity of our model, in that the EICANS agrees with the clinical gold standard, as defined by the ASTCT. The receiver operating curve shows substantial sensitivity and specificity in distinguishing mild and moderate as compared to severe ICANS, an important clinical threshold in decisions regarding diagnostic work-up and treatment.[1] The correlation between absolute error and maximum ICANS, as well as higher MAE among patients
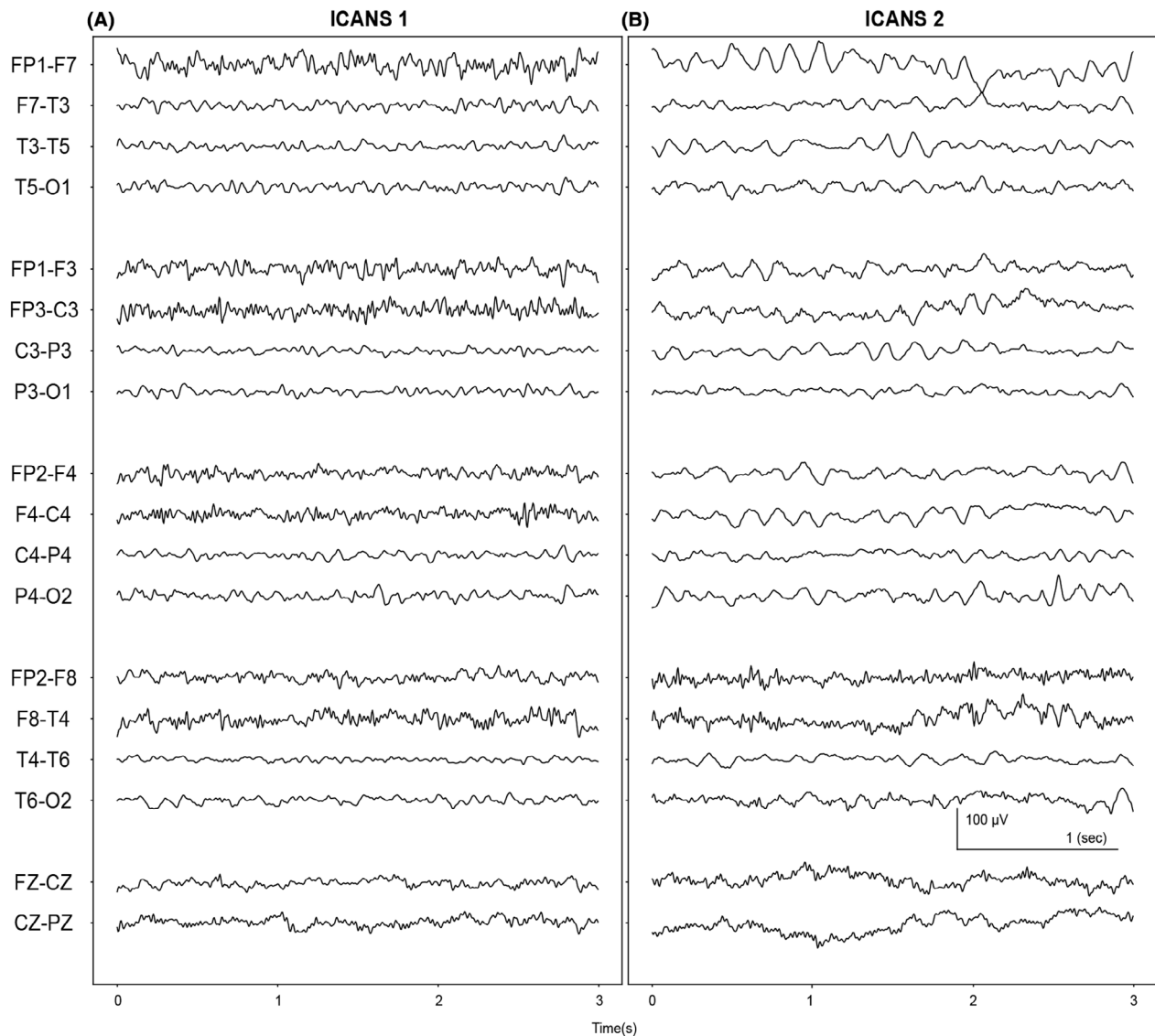
Figure 1. EEG shows qualitative differences associated with worsening ICANS. An example of an EEG (3 second window, bipolar montage) from the same patient exhibiting ICANS grade 1 (A) as compared to grade 2 (B). With more severe ICANS, this patient's EEG demonstrated more pronounced delta and theta slowing.

with more severe ICANS suggests less accuracy for more severe as compared to mild ICANS. This may suggest a tendency to underestimate ICANS in its most severe form, which likely arose due to fewer patients with ICANS of grade 3 or 4. Given that bedside detection of ICANS is more difficult with more mild ICANS, this greater accuracy with mild ICANS is likely more clinically relevant for patient care.

The EICANS model also exhibits construct validity, in that EICANS scores correlate with clinical variables associated with ICANS in prior studies.[8,10,26] Serum ferritin and dexamethasone use showed significant positive associations with EICANS. Minimum platelets during hospitalization exhibited a significant negative correlation with EICANS, which corroborates prior work indicating an association between lower platelet nadir and higher risk of ICANS, likely due to endothelial activation and resultant platelet consumption.[16,26] Other inflammatory markers, such as CRP and procalcitonin, showed positive associations, but may not have reached significance due to the bias of the dataset, in which only CAR-T patients who developed ICANS underwent EEG. Antipsychotic usage throughout hospitalization also correlated with EICANS, which likely arises from behavioral management
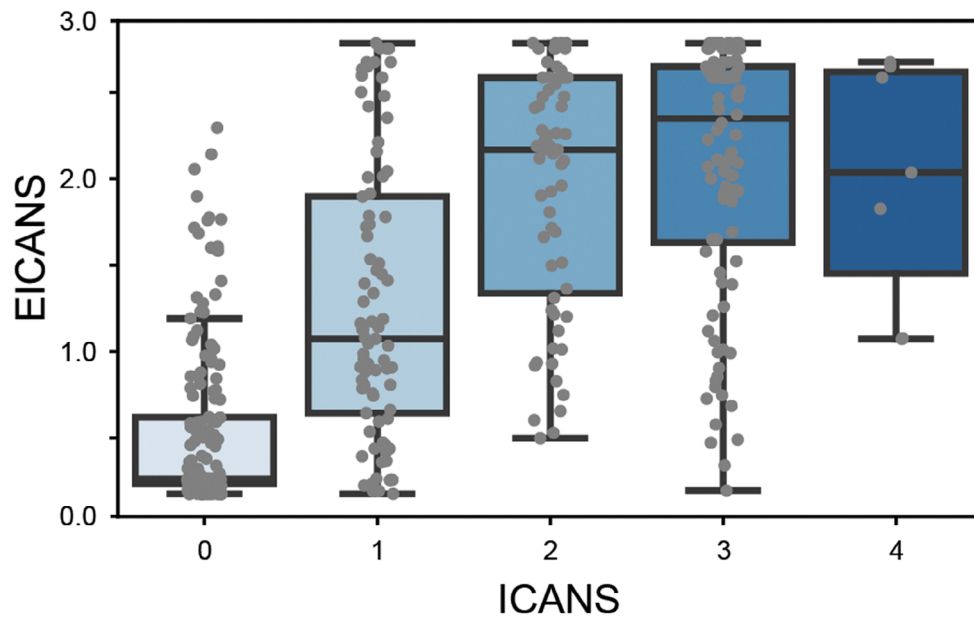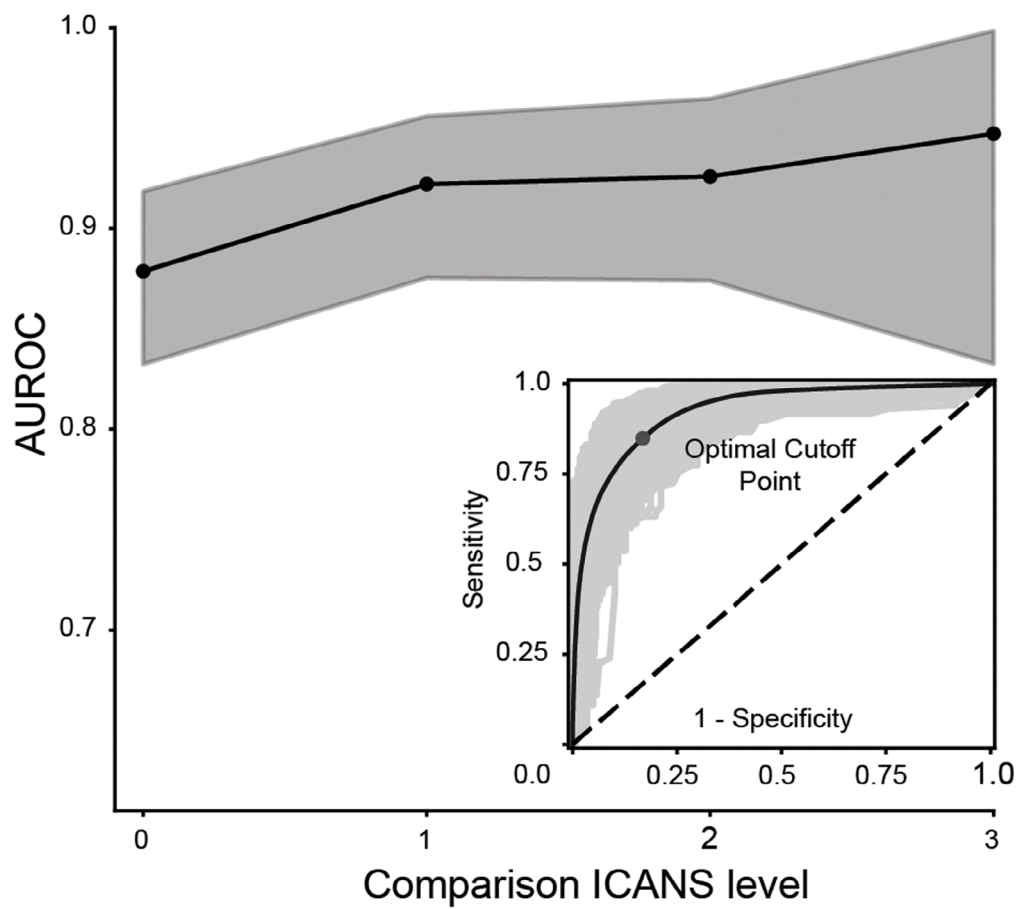
(A)

(B)

Figure 2. Criterion validity of EICANS. (A) Distribution of EICANS as compared to ICANS. The EICANS generated from the aggregated test folds shows a strong correlation with ICANS. The box plot (blue) shows the distribution of EICANS (*y*-axis) according to ICANS grade (0–4) (*x*-axis), with three quartile values (blue box) and values within 1.5 IQRs of the lower and upper quartile (black whiskers). The overlying points (gray) show each value of EICANS produced within the test folds of the model. (B) Area under the receiver operating curve for each comparison level of ICANS. The mean area under the receiver operating characteristic curve (AUROC) (*y*-axis) using EICANS to discriminate each level of ICANS (≤ x vs. > x for x = 0, 1, 2, & 3) (*x*-axis), with shading indicating the 95% confidence intervals for all bootstraps. The inset shows the mean and all bootstraps for the receiver operating characteristic (ROC) curve for ICANS ≤2 as compared to >2.

of ICANS. Tocilizumab did not show a clear relationship with EICANS, which may reflect the conflicting findings regarding the extent of tocilizumab's effect on ICANS when administered with concurrent steroids.[6]

We show the predictive validity of the model through the significant correlation between EICANS and duration of ICANS. Although there was a trend towards an association between severe ICANS and death at 1-year post discharge (Table 1), this association was not observed with EICANS or the maximum ICANS from days with EEG monitoring. This likely reflects the timing of EEG recording, in that EEGs were often collected during the first few days of ICANS, such that the maximum EICANS for the patient may not have matched the maximum ICANS over the entire hospitalization. Changing standards of ICANS care over the course of the study from 2016 to 2020, including clinical willingness to use dexamethasone and duration of admissions during early use of CAR-T cell therapy may have also affected our analysis of outcomes.

In addition to the EICANS grade, our model provides insights into EEG features indicative of ICANS, which expand upon prior qualitative EEG studies. Similar to delirium,[21,22,27] worsening ICANS has been correlated with slowing of the EEG.[6,8,10,18,28] However, the final model often did not incorporate low bandpass power frequencies (i.e., 0–2, 0–4) into EICANS, due to high collinearity with other retained predictive features. In contrast, spectral power in the theta range (4–6 Hz) was frequently retained with high median weight, indicating that this feature may discriminate among degrees of ICANS to a greater degree than the lower bandpass frequencies—a novel finding. The next most prominent features in our model, GPDs and GRDA, have also been shown to correlate with ICANS,[18,28] particularly with language dysfunction,[19] but prior work has not quantified these features or done so in an automated fashion. Lateralized periodic discharges (LPDs) and lateralized rhythmic delta activity (LRDA) did not contribute to the final EICANS model, likely due to low correlation with ICANS rather than collinearity with other features, which is in contrast to a prior study that had suggested that these features highly correlate with ICANS with focal findings.[28] This discrepancy may have arisen due to a low rate of LPD and LRDA within our dataset, or patients with LPD and

LRDA may reflect a subset clinical phenotype of ICANS not captured by the overall grade.

Among the negative predictive features of ICANS, the Non-IIIC feature is the most clinically intuitive result, in that this feature quantifies the absence of epileptiform patterns like seizures, rhythmic patterns, and periodic discharges. The Hurst exponent captures the long-term correlation structure ("memory") of a time series signal, with higher values indicating less volatility. Its role as a negative predictor in our model corroborates prior EEG findings in toxic-metabolic encephalopathy, in which high variability was associated with more severe delirium.[21]

This overlap of prominent EEG features between ICANS and toxic metabolic encephalopathy may indicate underlying similarities in pathophysiology. Although not entirely understood, the systemic inflammatory state induced by CAR-T is thought to cause endothelial activation and blood–brain barrier breakdown.[1,10] Similar changes in inflammation and blood brain barrier permeability have been demonstrated in sepsis and COVID, which can feature neurological changes similar to ICANS.[29,30] Due to these similarities, EEG may not be able to distinguish toxic-metabolic encephalopathy secondary to infection from encephalopathy related to CAR-T cell therapy without additional clinical data (i.e., culture and laboratory data). Timing may also inform interpretation; although there is no strict cutoff past which ICANS cannot be diagnosed, the median time to emergence is 5 days,[8] with clinical suspicion for other etiologies of encephalopathy increasing with greater time from infusion.

These findings possess notable limitations. The retrospective nature of the study meant that EEG data was only available from patients in whom providers clinically suspected ICANS. Due to the time interval between ordering an EEG and obtaining recordings, the dataset includes EEG samples from patients with an ICANS of 0, despite clinical suspicion for ICANS. This reflects the dynamic nature of ICANS, but also highlights the possibility of subtle encephalopathy not captured by the standardized clinical exam. It is possible that our control subjects, despite being matched for age and sex, could have some differences from EEG findings in patients undergoing CAR-T cell therapy who do not develop
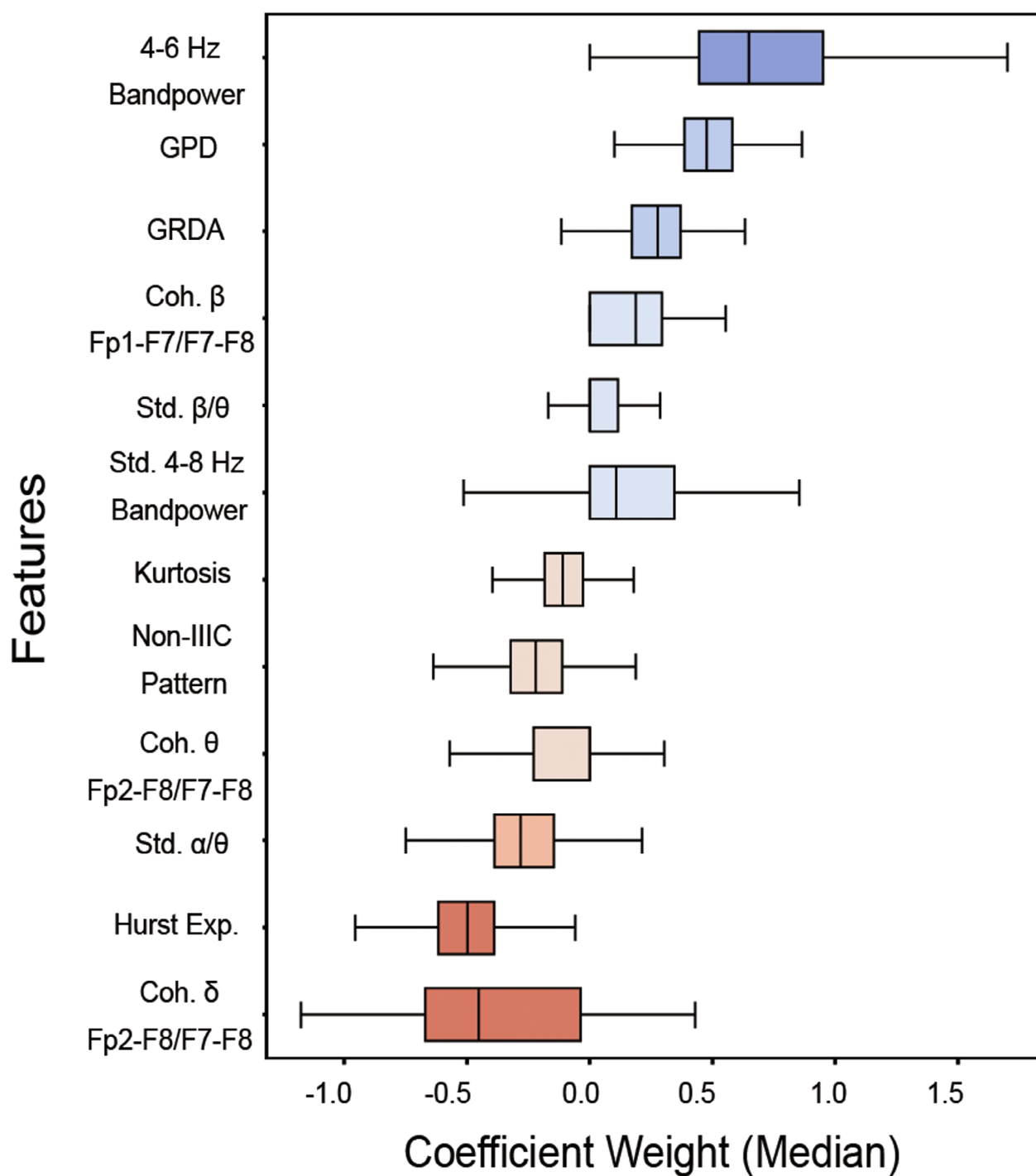
Figure 3. Median coefficient values of most highly weighted features. From the top 20 features most often selected by the Maximum Relevance-Minimum Redundancy algorithm for inclusion in the model, a boxplot shows all features (*y*-axis) with an absolute median coefficient value greater than 0.1, including negative (red) and positive (blue) weights. The boxes (*x*-axis) represent the three quartile values and the whiskers indicate the 1.5 IQRs of the lower and upper quartile.

ICANS. For patients with severe ICANS, sedation requirements certainly affect the EEG, but this likely did not affect the model given the minority of patients receiving propofol or continuous midazolam. Two patients were also excluded due to inability to lift sedation and obtain a clinical exam. CAR-T cell therapy patients often have
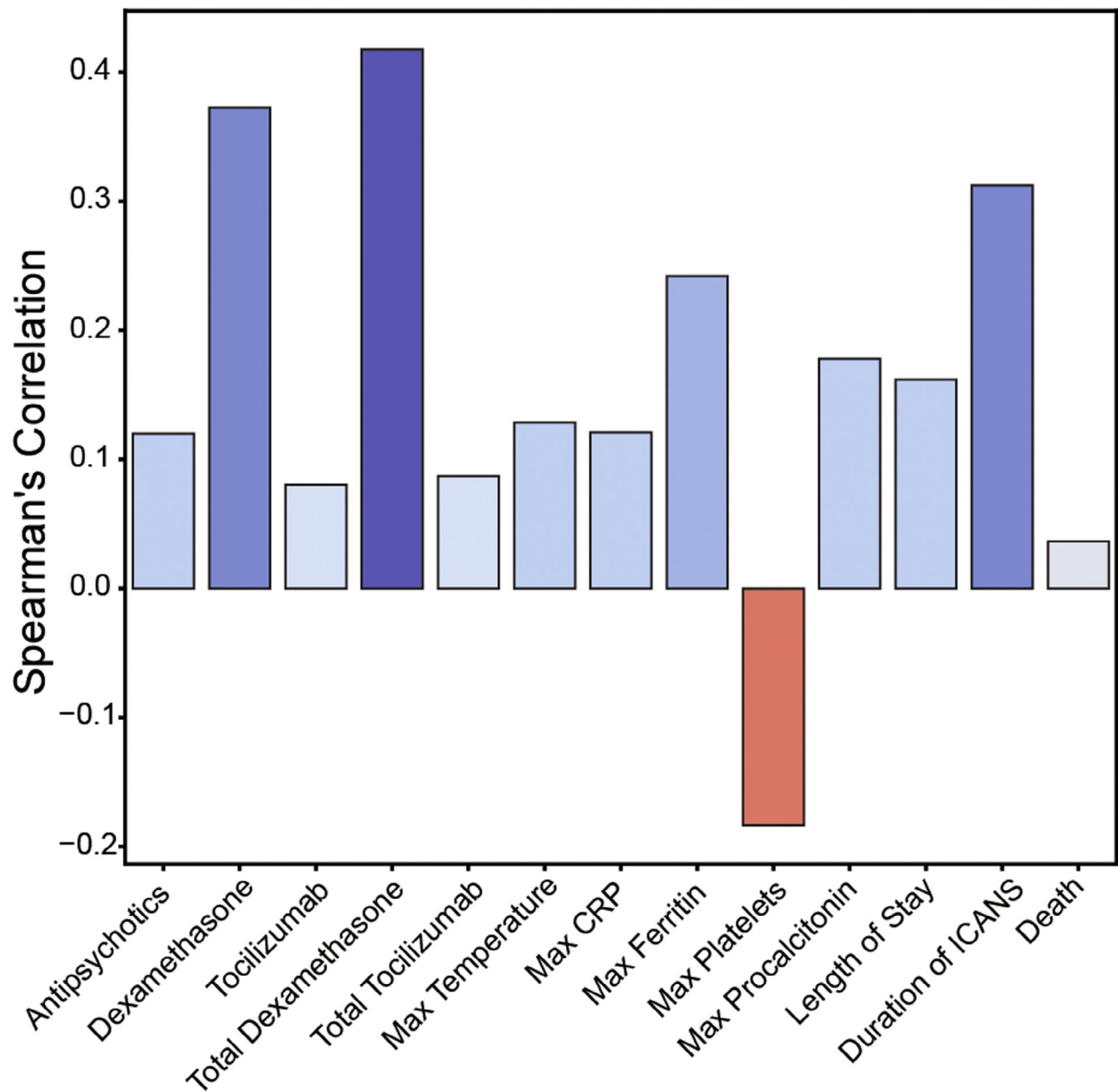
Figure 4. Spearman's correlations of EICANS and clinical variables. A bar plot shows the Spearman's correlations (*y*-axis) between the maximum EICANS per patient (*n* = 123) and clinical variables during each patient's hospitalization, including laboratory values, vitals, medications, and outcomes (death by 1 year post-discharge, length of stay, and duration of ICANS) (*blue* = positive coefficient value, *red* = negative coefficient value). Max refers to the maximum value of the laboratory or vital sign value during hospitalization up to and including the day that the EICANS was assessed.

medical comorbidities, such as febrile neutropenia, which may render their baseline EEG without ICANS different from those of our matched controls. The gold standard for the machine learning model relied upon a chart review-constructed ICANS grade, which could be liable to noise. Given the ICANS grade was assessed via clinical notes, the precise timing of examinations was not recorded. Due to this, the 9-min EEG sample selected from 24 h of continuous EEG monitoring may not have always occurred while the patient was experiencing the same degree of symptoms as reflected in the day's ICANS grade. Lastly, ICANS is a clinical syndrome defined by a subjective assessment, such that any model of ICANS will be trained with a potentially fallible gold standard.

However, through incorporation of objective data and training over large enough datasets, models such as these can begin to overcome this subjectivity to enhance reliability of diagnosis among providers. Despite these limitations, EICANS showed a strong, clinically significant correlation with ICANS, which substantiates the potential of this EEG-based method to dynamically assess ICANS severity.

Patients undergoing CAR-T cell therapy often have substantial medical comorbidities, such that their baseline EEGs may be abnormal in the absence of CNS involvement of their malignancy. Similar to patients with sepsis without neurological exam changes,[31,32] patients with CRS may also demonstrate subtle EEG changes, despite no apparent change in clinical exam, due to the profound inflammatory state. This sensitivity of EEG highlights its potential to predict ICANS. Future studies may advance these results through a prospective design with time-stamped clinical exams and baseline EEGs prior to CAR-T cell administration. This design would permit analysis of EEG predictors of ICANS prior to the emergence of ICANS, including duration and severity, as well as better assess the specificity of EEG changes during CRS alone versus ICANS. By obtaining EEG prior to CAR-T cell infusion, a more sophisticated model based on comparison to the patient's baseline could also be explored. Modification of the CNN-derived features to require only four frontal leads, similar to the majority of the current model's features, would also allow deployment of EICANS with simplified devices that do not require technician application.[33] Such algorithms could in turn facilitate the use of EEG and CAR-T cell therapy outside of specialized academic centers.

CAR-T cell therapy is a powerful oncologic treatment with the potential for wider adoption, given promising studies of its use in refractory autoimmune conditions[34] as well as malignancy. An objective method to assess ICANS such as EICANS may increase utilization of CAR-T cell therapy, enhance management of the condition, and reduce length of hospital stay.

Acknowledgments

Author Contributions

CAE and MBW contributed to the conception and design of the study, data analysis, and the drafting of manuscript and figures. HS contributed to data analysis and editing of the manuscript. PM, SQ, MS, DK, AJ, ZF, HD, CAJ, and DR contributed to data acquisition and management. MS, JJ, and WG assisted with data analysis. EYK, SSC, MJF, JWL, and JD contributed to design of study and drafting of the manuscript.

Conflict of Interest

MBW and SSC are co-founders of Beacon Biosignals, which played no role in this work. MJF received financial research support from Kite/Gilead and Novartis and consulting fees from Kite/Gilead, BMS, Novartis, and Iovance. DBR served on the scientific advisory board for Celgene/Bristol Meyers Squib. JD received research support from Novartis, consulting fees from Amgen, Blue Earth Diagnostics, and Syndax and royalties from contributing to UpToDate. CAJ received consulting fees from Kite/Gilead, Novartis, BMS/Celgene, Bluebird Bio, Epizyme, Instill-Bio, Lonza, Ipsen, Abintus-Bio, Daiichi-Sankyo and research funding from Kite/Gilead and Pfizer. Other authors report no disclosures relevant to the manuscript.

Data Availability Statement

The deidentified dataset and Python scripts utilized in the study is accessible upon request from the corresponding author.

References

1. Rubin DB, Vaitkevicius H. Neurological complications of cancer immunotherapy (CAR T cells). J Neurol Sci. 2021;424:117405. doi:10.1016/j.jns.2021.117405

2. Neelapu SS, Locke FL, Bartlett NL, et al. Axicabtagene Ciloleucel CAR T-cell therapy in refractory large B-cell lymphoma. N Engl J Med. 2017;377(26):2531-2544. doi:10.1056/NEJMoa1707447

3. Schuster SJ, Bishop MR, Tam CS, et al. Tisagenlecleucel in adult relapsed or refractory diffuse large B-cell lymphoma. N Engl J Med. 2019;380(1):45-56. doi:10.1056/NEJMoa1804980

4. Abramson JS, Palomba ML, Gordon LI, et al. Lisocabtagene maraleucel for patients with relapsed or refractory large B-cell lymphomas (TRANSCEND NHL 001): a multicentre seamless design study. Lancet. 2020;396 (10254):839-852. doi:10.1016/S0140-6736(20)31366-0

5. Munshi NC, Anderson LD, Shah N, et al. Idecabtagene vicleucel in relapsed and refractory multiple myeloma. N Engl J Med. 2021;384(8):705-716. doi:10.1056/NEJMoa2024850

6. Santomasso BD, Park JH, Salloum D, et al. Clinical and biological correlates of neurotoxicity associated with CAR T-cell therapy in patients with B-cell acute lymphoblastic leukemia. Cancer Discov. 2018;8(8):958-971. doi:10.1158/2159-8290.CD-17-1319

7. Lee DW, Santomasso BD, Locke FL, et al. ASTCT consensus grading for cytokine release syndrome and neurologic toxicity associated with immune effector cells. Biol Blood Marrow Transplant. 2019;25(4):625-638. doi:10.1016/j.bbmt.2018.12.758

8. Karschnia P, Jordan JT, Forst DA, et al. Clinical presentation, management, and biomarkers of neurotoxicity after adoptive immunotherapy with CAR T cells. Blood. 2019;133(20):2212-2221. doi:10.1182/blood-2018-12-893396

9. Torre M, Solomon IH, Sutherland CL, et al. Neuropathology of a case with fatal CAR T-cell-associated cerebral edema. J Neuropathol Exp Neurol. 2018;77 (10):877-882. doi:10.1093/jnen/nly064

10. Gust J, Hay KA, Hanafi LA, et al. Endothelial activation and blood-brain barrier disruption in neurotoxicity after adoptive immunotherapy with CD19 CAR-T cells. Cancer Discov. 2017;7(12):1404-1419. doi:10.1158/2159-8290.CD-17-0698

11. Locke FL, Neelapu SS, Bartlett NL, et al. Preliminary results of prophylactic tocilizumab after axicabtagenecilioleucel (axi-cel; KTE-C19) treatment for patients with refractory, aggressive non-hodgkin lymphoma (NHL). Blood. 2017;130(Supplement 1):1547. doi:10.1182/blood.V130.Suppl_1.1547.1547

12. Chen F, Teachey DT, Pequignot E, et al. Measuring IL-6 and sIL-6R in serum from patients treated with tocilizumab and/or siltuximab following CAR T cell therapy. J Immunol Methods. 2016;434:1-8. doi:10.1016/j.jim.2016.03.005

13. Davila ML, Riviere I, Wang X, et al. Efficacy and toxicity management of 19-28z CAR T cell therapy in B cell lymphoblastic leukemia. Sci Transl Med. 2014;6 (224):224ra25. doi:10.1126/scitranslmed.3008226

14. Guha-Thakurta N, Wierda WG. Cerebral edema secondary to chimeric antigen receptor T-cell immunotherapy. Neurology. 2018;91(18):843. doi:10.1212/WNL.0000000000006436

15. Gust J, Ishak GE. Chimeric antigen receptor T-cell neurotoxicity neuroimaging: more than meets the eye. AJNR Am J Neuroradiol. 2019;40(10):E50-E51. doi:10.3174/ajnr.A6184

16. Pennisi M, Sanchez-Escamilla M, Flynn JR, et al. Modified EASIX predicts severe cytokine release syndrome and neurotoxicity after chimeric antigen receptor T cells. Blood Adv. 2021;5(17):3397-3406. doi:10.1182/bloodadvances.2020003885

17. Rubin DB, Al Jarrah A, Li K, et al. Clinical predictors of neurotoxicity after chimeric antigen receptor T-cell

therapy. JAMA Neurol. 2020;77(12):1536-1542. doi:10.1001/jamaneurol.2020.2703

18. Herlopian A, Dietrich J, Abramson JS, Cole AJ, Westover MB. EEG findings in CAR T-cell therapy-related encephalopathy. Neurology. 2018;91(5):227-229. doi:10.1212/WNL.0000000000005910

19. Sokolov E, Karschnia P, Benjamin R, et al. Language dysfunction-associated EEG findings in patients with CAR-T related neurotoxicity. BMJ Neurol Open. 2020;2(1): e000054. doi:10.1136/bmjno-2020-000054

20. Sun H, Kimchi E, Akeju O, et al. Automated tracking of level of consciousness and delirium in critical illness using deep learning. NPJ Digit Med. 2019;2:89. doi:10.1038/s41746-019-0167-0

21. van Sleuwen M, Sun H, Eckhardt C, et al. Physiological assessment of delirium severity: the electroencephalographic confusion assessment method severity score (E-CAM-S). Crit Care Med. 2022;50(1): e11-e19. doi:10.1097/CCM.0000000000005224

22. Kimchi EY, Neelagiri A, Whitt W, et al. Clinical EEG slowing correlates with delirium severity and predicts poor clinical outcomes. Neurology. 2019;93(13):e1260-e1271. doi:10.1212/WNL.0000000000008164

23. Shoorangiz R, Weddell SJ, Jones RD. EEG-based machine learning: theory and applications. In: Thakor NV, ed. Handbook of Neuroengineering. Springer; 2020:1-39. doi:10.1007/978-981-15-2848-4_70-1

24. Ding C, Peng H, Peng H. Minimum redundancy feature selection from microarray gene expression data. J Bioinform Comput Biol. 2005;3:185-205. doi:10.1142/S0219720005001004

25. Burges C, Shaked T, Renshaw E, et al. Learning to rank using gradient descent. Proceedings of the 22nd International Conference on Machine Learning. ICML'05. Association for Computing Machinery; 2005:89-96. doi:10.1145/1102351.1102363

26. Hay KA, Hanafi LA, Li D, et al. Kinetics and biomarkers of severe cytokine release syndrome after CD19 chimeric antigen receptor–modified T-cell therapy. Blood. 2017;130 (21):2295-2306. doi:10.1182/blood-2017-06-793141

27. Tesh RA, Sun H, Jing J, et al. VE-CAM-S: visual EEG-based grading of delirium severity and associations with clinical outcomes. Crit Care Explor. 2022;4(1):e0611. doi:10.1097/CCE.0000000000000611

28. Beuchat I, Danish HH, Rubin DB, et al. EEG findings in CAR T-cell-associated neurotoxicity: clinical and radiological correlations. Neuro Oncol. 2022;24 (2):313-325. doi:10.1093/neuonc/noab174

29. Mazeraud A, Righy C, Bouchereau E, Benghanem S, Bozza FA, Sharshar T. Septic-associated encephalopathy: a comprehensive review. Neurotherapeutics. 2020;17 (2):392-403. doi:10.1007/s13311-020-00862-1

30. Lin L, Al-Faraj A, Ayub N, et al. Electroencephalographic abnormalities are common in COVID-19 and are

associated with outcomes. Ann Neurol. 2021;89 (5):872-883. doi:10.1002/ana.26060

31. Velissaris D, Pantzaris ND, Skroumpelou A, et al. Electroencephalographic abnormalities in sepsis patients in correlation to the calculated prognostic scores: a case series. J Transl Int Med. 2018;6(4):176-180. doi:10.2478/jtim-2018-0032

32. Pantzaris ND, Platanaki C, Tsiotsios K, Koniari I, Velissaris D. The use of electroencephalography in patients with sepsis: a review of the literature. J Transl Int Med. 2021;9(1):12-16. doi:10.2478/jtim-2021-0007

33. Singla S, Garcia GE, Rovenolt GE, et al. Detecting seizures and Epileptiform abnormalities in acute brain injury. Curr Neurol Neurosci Rep. 2020;20(9):42. doi:10.1007/s11910-020-01060-4

34. Mougiakakos D, Krönke G, Völkl S, et al. CD19-targeted CAR T cells in refractory systemic lupus erythematosus. N Engl J Med. 2021;385(6):567-569. doi:10.1056/NEJMc2107725